# Workshop on explainable AI in automated driving: a user-centered interaction approach

**Conference Paper** · September 2019

DOI: 10.1145/3349263.3350762



**9 authors**, including:

Quentin Meteier
University of Applied Sciences and Arts Western Switzerland
**5** PUBLICATIONS **5** CITATIONS

SEE PROFILE

Marine Capallera
School of Management Fribourg
**7** PUBLICATIONS **6** CITATIONS

SEE PROFILE

# Workshop on Explainable AI in Automated Driving: A User-Centered Interaction Approach

**Quentin Meteier**
**Marine Capallera**
**Leonardo Angelini**
**Elena Mugellini**
**Omar Abou Khaled**
HumanTech Institute, HES-SO // University of Applied Sciences Western Switzerland
quentin.meteier@hes-so.ch
marine.capallera@hes-so.ch
leonardo.angelini@hes-so.ch
elena.mugellini@hes-so.ch
omar.aboukhaled@hes-so.ch

**Stefano Carrino**
Haute Ecole Arc Ingénierie, HES-SO
Neuchâtel, Switzerland
stefano.carrino@hes-so.ch

**Emmanuel De Salis**
Haute Ecole Arc Ingénierie, HES-SO
Saint-Imier, Switzerland
emmanuel.desalis@hes-so.ch

**Stéphane Galland**
CIAD, Univ. Bourgogne Franche-Comté, UTBM
Belfort, France
stephane.galland@utbm.fr

**Susanne Boll**
University of Oldenburg
Oldenburg, Germany
susanne.boll@uol.de


*AutomotiveUI '19 Adjunct* , September 21–25, 2019, Utrecht, Netherlands

ACM ISBN 978-1-4503-6920-6/19/09 .
https://doi.org/10.1145/3349263.3350762

## Abstract
With the increasing use of automation, users tend to delegate more tasks to the machines. Such complex systems are usually developed with "black box" Artificial Intelligence (AI), which makes these systems difficult to understand for the user. This assumption is particularly true in the field of automated driving since the level of automation is constantly increasing via the use of state-of-the-art AI solutions. We believe it is important to investigate the field of Explainable AI (XAI) in the context of automated driving since interpretability and transparency are key factors for increasing trust and security. In this workshop, we aim at gathering researchers and industry practitioners from different fields to brainstorm about XAI with a special focus on human-vehicle interaction. Questions like "what kind of explanation do we need", "which is the best trade-off between performance and explainability" and "how granular should the explanations be" will be addressed in this workshop.

## Author Keywords
Automated driving; explainable artificial intelligence; trust in automation; user interface design; vehicle environment

## CCS Concepts
**•Human-centered computing → Human computer interaction (HCI); •Computing methodologies → Artificial intelligence;**

## Introduction

Today's vehicles are helping the driver in his or her driving task and offer safer driving. Indeed, Advanced Driver-Assistance Systems (ADAS) and more particularly SAE level 2 cars [6], allow the car to maintain its speed, its safety distance from the front car, but also its position in the center of its lane. However, the driver must keep his/her hands on the wheel and constantly monitor the surrounding environment. That is why the increase in the level of automation through the use of artificial intelligence, including machine learning and deep learning algorithms, allows the vehicle to drive more and more autonomously without requiring driver supervision (e.g., level 3 and 4 SAE car, [6]). However, the use of AI does not yet seem to be suitable for a majority of people. Indeed, drivers (and pedestrians) do not yet want to have complete confidence in a self-driving vehicle by delegating all driving tasks to it. In addition, the driver may not understand the vehicle's intentions and decisions because the autonomous system would behave differently from his/her expectations. In the same way, pedestrians could be surprised by vehicle's behaviour because they wouldn't perform eye contact with the driver anymore. In the traffic, eye contact between drivers and between drivers and other road users (such as cyclists and pedestrians) plays a critical role. Often, cyclists entering the roadway or pedestrians who want to cross the street use eye contact to ensure that an approaching driver sees them or that the driver is not distracted. If the driver returns the eye contact, they can assume that the driver will act accordingly ([3], [8]). In some fully automated driving situations (or partially automated driving with driver supervision), if the driver was driving, he/she would have avoided the obstacle, but in the case of fully automated driving, the vehicle would prefer to brake [9]. Because AI can confront driver (and also pedestrians) expectations, it is important that AIs communicate their status and intentions to users. Another factor that needs to be taken into account is the handling of take-over requests. Information given to the driver has to be quickly understood, but also easily and intuitively handled by the driver [5].

In order to better understand how AI works and build trust in the decisions made by AIs, new techniques in the field are emerging that are referred to as Explainable AI (XAI). These techniques are intended to "produce more explainable models, while maintaining a high level of learning performance (prediction accuracy) and enable human users to understand, appropriately trust, and effectively manage the emerging generation of artificially intelligent partners" [4]. As we can see, the main purposes of this transparency are to:

- understand the functioning of algorithms and AIs in order to optimize their design and architecture, their features but also to understand and interpret the results
- increase human confidence in systems
- increase and improve cooperation between agents

As shown by [11], providing appropriate explanations to the user increases the user's confidence in the system and thus allows for better human-AI collaboration. Prashan Madumal et al. [7] have shown that it is possible to establish an appropriate interaction protocol and proposed one for the socio-cognitive process of explanation in the domain of dialogue and conversation between humans and AIs. XAI is the subject of extensive research in the field of algorithmic trading, medical diagnostics and, of course, autonomous vehicles. Cysneiros et al. [2] investigated how transparency in self-driving cars affect driver trust and acceptance; the authors investigated how to pursue the elicitation and modeling of transparency. Moreover, transparency and trust

could work collaboratively to help build public and regulatory confidence in Automated Systems and have a vital role in the development of such systems.

In the field of autonomous driving, focusing on XAIs would increase drivers and pedestrians trust and acceptance and help Human-Vehicle Interaction (HVI) designers and engineers to provide more efficient interactions.

**Topics of Interest**
The workshop major topics comprehend, but are not limited to:
i) Theoretical Approaches for Explainability
ii) Explainable and Transparent AI
iii) Human-centered XAI
iv) Interactive XAI
v) Interpretable Machine Learning
vi) Human-AI interaction and collaboration
vii) Multimodal Interaction and Interaction Design
viii) Decision making in complex and real-time situations
ix) Fair, Accountable and Transparent algorithms to avoid AI bias

## Objectives

We seek to collect participants' contributions in the field of XAI for Automated Vehicles with short position papers. The list of topics comprehend, but is not limited to propositions on the left.

During the second part of the workshop, participants will engage in two practical applications of XAI in the automotive field: i.e., XAI interfaces for the driver and for pedestrians. To this purpose discussions will be focused on

- Identification of important information that might (or should) be provided respectively to drivers and pedestrians, in particular referring to the different levels of situation-awareness related to agent transparency [1]
- Explore and design with participants UI solutions using various interaction modalities in order to convey information defined above [10]

In order to center discussion around concrete conditionally automated driving situations, we will offer participants the opportunity to experience a pre-selected choice of critical scenarios using VR simulation (from the driver's and pedestrian's point of view) to better understand the issues.

## Workshop plan

*Workshop preparation and outreach*
A workshop website[1] will be published before the call for papers. Organizers will then post the call for participation in the workshop website. The organizers will also publicize the call in their home organizations and among their peers. Short papers between 3 and 6 pages long in the SIGCHI Extended Abstract template will be accepted (not including references and auxiliary information). They will be reviewed by the program committee based on their relevance to the topic and their pertinence to bring novel ideas in the field of XAI or HMI. The workshop will be open to the attendees with an accepted paper, but also for people from car manufacturing industry and academic researchers that would be simply interested by workshop topics. We plan to contact a publisher in order to include the accepted papers in a special issue of a scientific journal. We expect to attract 20-30 participants.

*Workshop schedule*
The workshop will follow a half-day format (4 hours). A tentative plan for the workshop schedule is proposed in Table 1. To start the workshop, an introductory session will be given by the organizers in order to detail the organisation and explain the objectives and expected results of this workshop. Attendees will have 2-3 minutes each to make a pitch about topics of interests proposed in the margin. Then, a session of 45 minutes will be allocated for participants that have an accepted paper in order to present their work to other attendees. Then, organizers will split participants into 4 groups of 5 to 7 people. A first session of work group will take place. It consists in putting participants in a realistic driving situation in virtual reality. Two virtual reality simulation stations will be installed in the room. This idea is to propose a realistic, attractive and immersive driving

[1] https://sites.google.com/view/explainableai-autoui19/home

**Important dates**
*Deadline for submissions*:
September 04, 2019
*Response to authors*:
September 15, 2019
*Camera ready submission deadline*:
September 20, 2019
*Workshop day*:
September 22, 2019

| Session | Dur. |
|---|---|
| Introductory session | 15 min |
| Short papers presentation | 45 min |
| Instructions | 5 min |
| Presentation of driving scenarios | 5 min |
| Brainstorming | 50 min |
| Break | 15 min |
| Creativity session | 60 min |
| Presentations and discussion | 30 min |
| Closing session | 15 min |

**Table 1:** Schedule

scenario to the attendees in order to stimulate their creativity. Each station will show driving scenarios with a different point of view : one from the driver's seat and one from a pedestrian point of view. Each point of view will be experienced by two groups.

While an attendee is watching the scene with the VR headset, other participants will have the opportunity to watch the same scenario shown through a video that will be looped and projected on a large screen. According to the driving scenarios they will experience, each group will have to brainstorm about what information should be conveyed to the driver/pedestrian about the state of the autonomous vehicle and what modality should be used to convey the information. After the break, attendees will do a creativity session. Based on the ideas that result from the brainstorming, each group will have to create a prototype or a mockup to answer the problem shown by the driving scenario. It is expected that the prototype/mockup give insights about what information to the driver and how to convey the information, possibly using different modalities. To support this phase, tools for sketching and rapid prototyping will be proposed. Finally, each group will have 10 minutes to present their prototype to the other attendees.

*Expected outcomes*
The expected outcome for the first session of the workshop is a summary of actual needs, challenges and opportunities for human-centered XAI in Automated Driving. The expected outcome for the second and third session of the workshop (brainstorming and creativity) are some practical insights for more human-centered approaches to XAI and examples of novel ways to provide human-centered explanations. In a longer term, we expect the creation of an interdisciplinary community interested in studying the human's role in explainability approaches in Automated Driving with adepts from both academia and industry community. We hope that in the future the proper design of human-AI interaction and collaboration in the automotive community will improve user experience and trust in car while increasing safety on the roads.

## Short Biography of the Organizers

**Quentin Meteier** is a PhD student at HumanTech Institute. He is working on conditionally automated driving and he is focusing more specifically on using psychophysiological signals of the driver in order to build a model describing the driver's state while driving a conditionally automated vehicle.

**Marine Capallera** is a PhD student at HumanTech Institute. She is working on conditionally automated driving and she is focusing more specifically on multimodal Human-Vehicle Interaction model for supervision.

**Leonardo Angelini** is a HCI post-doctoral researcher at the HumanTech Institute and Lecturer at HES-SO (teaching HCI and Machine Learning). He has run several workshops in interaction design, including workshops on tangible interaction with IoT at Ubicomp'16 and CHI'18, workshops on full-body and multisensory interaction at TEI'16 and Ubicomp'16, and a workshop on wearable computing at Automotive UI'14.

**Elena Mugellini** is head of the HumanTech Institute and Professor at HES-SO (teaching HCI and Machine Learning). She has run several workshops in interaction design, including workshops on tangible interaction at TEI'15 and CHI'18, wearable computing at Ubicomp'13 and Ubicomp'14, conversational agents at Ubicomp'18.

**Omar Abou Khaled** is Professor at the University of Applied Sciences and Arts Western Switzerland. His research

fields are Human-Computer Interaction and Wearable and Ubiquitous computing.

**Stefano Carrino** is Professor in computer science at the University of Applied Sciences of Western Switzerland in Neuchâtel (HES-SO). He is a member of the Data Analytics research group where he leads several applied projects related to AI and Machine Learning. His research domains are in the area of Artificial Intelligence, Human-Computer Interaction, Industry 4.0, and Serious Games and Gamification for Health.

**Emmanuel de Salis** is a PhD student at the University of Applied Sciences of Western Switzerland in Neuchâtel (HES-SO), and a member of its Data Analytics Research group. He is working on conditionally automated driving and more specifically on how to support the driver with AI when he needs to perform a take-over of the vehicle control.

**Stéphane Galland** is the deputy director of the Distributed Knowledge and Artificial Intelligence Laboratory and full Professor at UTBM. He is head of the Software Engineering section of the UTBM Computer Science Department. His research domains are in the fields of Multi-agent Systems, Agent-based Software Engineering, Agent-based Simulation, applied to complex systems and cyber-physical systems.

**Susanne Boll** is a full professor for Media Informatics and Multimedia Systems at the University of Oldenburg and a member of the board of the OFFIS-Institute for Information Technology. Her research interests lie in the field of human computer interaction (HCI), specifically in the area of pervasive user interfaces, mobile, and wearable interactive systems. Her scientific results have been published in competitive peer-reviewed international conferences such as CHI, MobileHCI, AutomotiveUI, DIS, ACM Multimedia and IDC, as well as internationally recognized journals.

## REFERENCES


1. Jessie Y. C. Chen, Shan G. Lakhmani, Kimberly Stowers, Anthony R. Selkowitz, Julia L. Wright, and Michael Barnes. 2018. Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theoretical Issues in Ergonomics Science* 19, 3 (2018), 259–282. DOI: http://dx.doi.org/10.1080/1463922X.2017.1315750
2. L. M. Cysneiros, M. Raffi, and J. C. Sampaio do Prado Leite. 2018. Software Transparency as a Key Requirement for Self-Driving Cars. In *2018 IEEE 26th International Requirements Engineering Conference (RE)*. 382–387. DOI: http://dx.doi.org/10.1109/RE.2018.00-21
3. B. Färber. 2016. *Communication and communication problems between autonomous vehicles and human drivers*. Springer, Berlin, Heidelberg. 125–144 pages. DOI:http://dx.doi.org/https://doi.org/10.1007/978-3-662-48847-8_7
4. David Gunning. 2017. Explainable Artificial Intelligence (XAI). Defense Advanced Research Projects Agency (DARPA). Accessed on 2019-07-23. (2017). http://www.darpa.mil/program/explainable-artificial-intelligence
5. Monika Hengstler, Ellen Enkel, and Selina Duelli. 2016. Applied artificial intelligence and trust - The case of autonomous vehicles and medical assistance devices. *Technological Forecasting and Social Change* 105 (2016), 105 – 120. DOI:http://dx.doi.org/https://doi.org/10.1016/j.techfore.2015.12.014
6. SAE International. 2018. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. (2018).
7. Prashan Madumal, Tim Miller, Liz Sonenberg, and Frank Vetere. 2019. A Grounded Interaction Protocol for Explainable Artificial Intelligence. *CoRR* abs/1903.02409 (2019). http://arxiv.org/abs/1903.02409
8. Sabrina Schmidt and Berthold Faerber. 2009. Pedestrians at the kerb – Recognising the action intentions of humans. *Transportation Research Part F-traffic Psychology and Behaviour - TRANSP RES PT F-TRAFFIC PSYCH* 12 (07 2009), 300–310. DOI: http://dx.doi.org/10.1016/j.trf.2009.02.003
9. Vindhya Venkatraman, John D. Lee, and Chris W. Schwarz. 2016. Steer or Brake?: Modeling Drivers' Collision-Avoidance Behavior by Using Perceptual Cues. *Transportation Research Record* 2602, 1 (2016), 97–103. DOI:http://dx.doi.org/10.3141/2602-12
10. Himanshu Verma, Florian Evéquoz, Guillaume Pythoud, Grace Eden, and Denis Lalanne. 2019. Engaging Pedestrians in Designing Interactions with Autonomous Vehicles. DOI: http://dx.doi.org/10.1145/3290607.3312864
11. Serena Villata, Guido Boella, Dov M. Gabbay, and Leendert van der Torre. 2013. A socio-cognitive model of trust using argumentation theory. *International Journal of Approximate Reasoning* 54, 4 (2013), 541 – 559. DOI:http://dx.doi.org/https://doi.org/10.1016/j.ijar.2012.09.001 Eleventh European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2011).